



①9 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

⑫ Patentschrift  
⑩ DE 195 28 425 C 1

⑤1 Int. Cl.<sup>6</sup>:  
H 04 N 13/00

⑳ Aktenzeichen: 195 28 425.9-31  
㉑ Anmeldetag: 2. 8. 95  
㉒ Offenlegungstag: —  
㉓ Veröffentlichungstag  
der Patenterteilung: 15. 5. 96

DE 195 28 425 C 1

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

㉔ Patentinhaber:  
Siemens AG, 80333 München, DE

㉕ Erfinder:  
Riegel, Thomas, Dipl.-Inform., 81739 München, DE;  
Ziegler, Manfred, Dipl.-Inform., 81739 München, DE

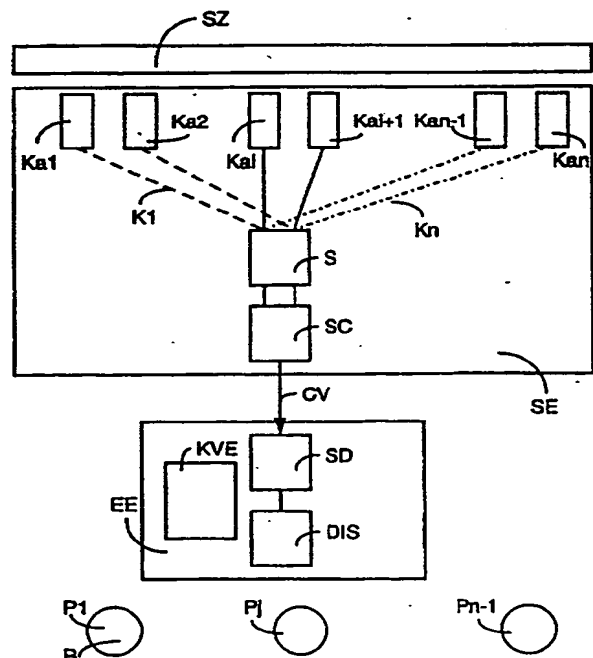
㉖ Für die Beurteilung der Patentfähigkeit  
in Betracht gezogene Druckschriften:

DE 40 27 471 C2  
EP 06 50 301 A2

㉗ Anordnung und Verfahren zur automatischen Auswahl zweier Kamerasignale aus einer Vielzahl von Kamerasignalen durch Bestimmung der Kopfposition eines Betrachters

㉘ Durch eine Kopfverfolgungseinheit (KVE) wird eine Kopfposition (Pj) eines Betrachters (B) bestimmt, und es werden zwei Kamerasignale (Ki; Ki+x) ausgewählt und von einer Sendeeinheit (SE) angefordert. Durch eine Schaltereinheit (S) werden die angeforderten Kameras (Kai; Kai+x) an eine stereoskopische Codiereinheit (SC) durchgeschaltet und dort zu einem codierten Videodatenstrom (CV) codiert. Dieser codierte Videodatenstrom (CV) wird an eine Empfangereinheit (EE) übertragen, dort decodiert und auf einem stereoskopischen Bildschirm (DIS) dem Betrachter dargestellt.

Damit wird eine sehr einfache automatisierte Auswahl von zwei benötigten Kamerasignalen (Ki; Ki+x) aus einer Vielzahl von Kamerasignalen (Ki) erreicht.



DE 195 28 425 C 1

Bei der Übertragung von sogenannten Multiview-Stereobildsequenzen, bei denen mehr als zwei Kameras eine Szene aufnehmen, existieren mehrere Probleme bei der Übertragung der Kamerasignale, die die aufgenommene Szene beschreiben. Bei den Multiview-Stereobildsequenzen benötigt der Empfänger zwei Kamerasignale, die dieselbe Szene aus verschiedenen Ansichten beschreiben. Durch die bestehende Disparität der Bilder der aufgenommenen Kamerasignale, das heißt durch die örtliche Verschiebung der entsprechenden Bildpunkte zwischen den Bildern der zwei Kamerasignale, die bei einem Empfänger auf einem stereoskopischen Bildschirm dargestellt werden, wird für den Betrachter ein Tiefeneffekt erzielt, das heißt, der Betrachter hat den Eindruck, ein dreidimensionales Bild zu betrachten.

Wenn eine Szene von mehreren Kameras aufgenommen wird, und alle Kamerasignale von allen Kameras an den Empfänger übertragen würden und die Auswahl der zwei benötigten Kamerasignale beim Empfänger stattfände, würde eine sehr große Menge an Übertragungskapazität benötigt, die unnötig viel Redundanz beinhalten würde.

Da in der Empfängereinheit nur zwei, allerdings die für den Betrachter "richtigen" Kamerasignale benötigt werden ist es wünschenswert, wirklich nur die benötigten zwei Kamerasignale zu übertragen. Weiterhin ist es wünschenswert, die Ansichten der Szene variieren zu können, ohne dafür größeren Aufwand treiben zu müssen.

Im übrigen ist es aus der EP 0650301 A2 bekannt, bei der stereoskopischen Bildaufnahme mehr als zwei Kameras einzusetzen. Des weiteren betrifft die DE 40 27 471 C2 eine von der sensorisch erfaßten Position des Betrachters abhängige räumliche Darstellung von Bildern.

Somit liegt der Erfindung das Problem zugrunde, eine Anordnung und ein Verfahren anzugeben, mit dem auf einfache Weise zwei Kamerasignale aus einer Vielzahl von n Kameras erzeugten Kamerasignalen, die eine Szene aus verschiedenen Ansichten aufnehmen, automatisch von einem Betrachter der Szene, der sich beim Empfänger befindet, ausgewählt werden.

Dieses Problem wird durch die Anordnung gemäß Patentanspruch 1 und dem Verfahren gemäß Patentanspruch 3 gelöst.

Ein Vorteil der erfindungsgemäßen Anordnung liegt vor allem darin, daß die Anordnung sehr einfach ist und somit nur geringfügige Erweiterungen gegenüber Anordnungen notwendig sind, die ohnehin benötigt werden, um einen stereoskopischen Videodatenstrom auf Basis zweier fest angeschlossener Kameras, die eine Szene aufnehmen, zu realisieren.

Auch ist das erfindungsgemäße Verfahren durch seine Einfachheit geprägt, was zu einer sehr schnellen Durchführung des Verfahrens führt, wodurch eine Echtzeitrealisierung des Verfahrens möglich ist, ohne störende Verzögerungseffekte für den Betrachter bei der Auswahl einer anderen Ansicht einer betrachteten Szene.

Die Verwendung des Verfahrens gemäß Patentanspruch 4 als eine Ausnahmebehandlung in einem Fehlerfall, also als eine sogenannte Fallback-Lösung, bei der Verwendung von modellbasierten Codierungsverfahren zur Codierung eines Videodatenstroms, erhöht die Verlässlichkeit der modellbasierten Codierungsverfahren erheblich.

Weiterbildungen der Erfindung ergaben sich aus den

abhängigen Ansprüchen.

Ein bevorzugtes Ausführungsbeispiel der Erfindung und eine vorteilhafte Verwendung des erfindungsgemäßen Verfahrens sind in den Zeichnungen dargestellt und werden im folgenden näher beschrieben.

Es zeigen

Fig. 1 eine Skizze, die die Anordnung mit einer Empfängereinheit und einer Sendeeinheit und der Vielzahl von Kameras, die anhand der Kopfposition des Betrachters ausgewählt werden, darstellt;

Fig. 2 ein Ablaufdiagramm, das das erfindungsgemäße Verfahren gemäß Patentanspruch 3 beschreibt;

Fig. 3 ein Ablaufdiagramm, das die Verwendung des erfindungsgemäßen Verfahrens als Fallback-Lösung bei einer modellbasierten Codierung darstellt.

Anhand der Fig. 1 bis 3 wird die erfindungsgemäße Anordnung, das erfindungsgemäße Verfahren sowie eine Verwendung des erfindungsgemäßen Verfahrens weiter erläutert.

In Fig. 1 ist die erfindungsgemäße Anordnung beschrieben, die mindestens folgende Komponenten aufweist:

— eine Sendeeinheit SE, die mindestens folgende Komponenten aufweist:

— eine Vielzahl von Kameras Kai, die eine Szene aus mehreren Ansichten, das heißt von verschiedenen Perspektiven, aufnehmen; ein erster Index i, der jede Kamera eindeutig identifiziert, ist eine Zahl, die in dem Bereich zwischen 1 und n liegt, wobei die Werte 1 und n in dem Bereich enthalten sind;

— eine Schaltereinheit S, mit n Eingängen, wobei jeweils ein Eingang der Schaltereinheit S mit einer Kamera Kai gekoppelt ist, und zwei Ausgängen,

— eine stereoskopische Codiereinheit SC zur Codierung eines zu sendenden codierten Videodatenstroms CV;

— eine Empfängereinheit EE, die mindestens folgende Komponenten aufweist:

— eine stereoskopische Decodiereinheit SD, in der die Decodierung des empfangenen codierten Videodatenstromes CV durchgeführt wird,

— eine Kopfverfolgungseinheit KVE, in der die Kopfposition Pj eines Betrachters B bestimmt wird, in der aus den bestimmten Koordinaten der Kopfposition die benötigten zwei Ansichten, das heißt zwei benötigte Kamerasignale Ki, wobei der erste Index i auch das Kamerasignal, das von der i-ten Kamera Kai zur Verfügung gestellt wird, bezeichnet, ausgewählt werden; ein zweiter Index j, der jede Kopfposition Pj des Betrachters B eindeutig identifiziert, ist eine Zahl, die in dem Bereich zwischen 1 und n-1 liegt, wobei die Werte 1 und n-1 in dem Bereich enthalten sind,

— ein stereoskopischer Bildschirm DIS, der einen decodierten Videodatenstrom DCV dem Betrachter B in der Weise darstellt, daß für den Betrachter B ein stereoskopischer Effekt, das heißt ein Tiefeneindruck bei Betrachtung des Bildes, entsteht.

Die Kameras Kai nehmen eine Szene SZ auf. Die Kameras Kai sind in beliebiger Weise angeordnet.

Dies ist eine Voraussetzung dafür, daß bei einer belie-

bigen Auswahl zweier Kameras Kai und Kai + x, wobei eine Zahl x eine in dem Bereich 1 bis n-1 liegt, ein stereoskopischer Effekt für den Betrachter B erreicht werden kann.

Die Sendeeinheit SE weist mindestens die n Kameras Kai, die Schaltereinheit S und die stereoskopische Codiereinheit SC auf.

Die Schaltereinheit S ist mit den zwei Ausgängen mit der stereoskopischen Codiereinheit SC gekoppelt. Der Aufbau einer stereoskopischen Codiereinheit SC ist dem Fachmann bekannt und beispielsweise in (R. ter Horst, A Demonstrator for Coding and Transmission of Stereoscopic Video Signals, Proceedings of 4-th European Workshop on Three-Dimensional Television, Rom, 20.-21. Oktober 1993, S. 273-280, 1993) beschrieben.

Der zu sendende, in der stereoskopischen Codiereinheit SC codierte Videodatenstrom CV wird an die Empfängerseinheit EE übertragen. Jeder der n Eingänge der Schaltereinheit S sind mit jeweils einer Kamera Kai gekoppelt.

Entsprechend einer Kameraanforderungsnachricht KAN, die im weiteren beschrieben wird, werden in der Schaltereinheit S die durch die Kameraanforderungsnachricht KAN angeforderten Kameras Kai und Kai + 1 ausgewählt, so daß jeweils die angeforderten Kamerasignale Ki und Ki + 1 der Kameras Kai und Kai + 1 codiert und übertragen werden.

Somit werden immer nur die angeforderten Kamerasignale ausgewählt und an die stereoskopische Codiereinheit SC durchgeschaltet 3, wo sie zu dem codierten Videodatenstrom CV codiert werden 4 und übertragen werden an die Empfängerseinheit EE.

In der Empfängerseinheit EE wird der codierte Videodatenstrom CV empfangen und decodiert 5. Anschließend wird der decodierte Videodatenstrom auf dem stereoskopischen Bildschirm DIS dem Betrachter B dargestellt.

Die Empfängerseinheit EE weist mindestens die stereoskopische Decodiereinheit SD, die Kopfverfolgungseinheit KVE sowie den stereoskopischen Bildschirm DIS auf 6.

Die stereoskopische Decodiereinheit SD ist mit dem stereoskopischen Bildschirm DIS gekoppelt. In der Kopfverfolgungseinheit KVE wird die Kopfposition des Betrachters B bestimmt, indem mit einer weiteren Kamera der Kopf des Betrachters B aufgenommen wird, erkannt wird und die räumliche Lage des Kopfes des Betrachters B, also die Kopfposition Pj des Betrachters B bestimmt wird.

Anhand der Kopfposition Pj, wobei der zweite Index j in dem Bereich zwischen 1 und n-1 liegt, werden die der Kopfposition Pj zugeordneten Kamerasignale Ki und Ki + 1 ausgewählt und von der Kopfverfolgungseinheit KVE angefordert.

In der Kopfverfolgungseinheit KVE wird die Kopfposition Pj des Betrachters B bestimmt 1 und die horizontalen Bewegungen des Betrachters B im Vergleich zum Bildschirm verfolgt. Hierzu wird mindestens eine Kamera verwendet, die den Kopf des Betrachters B aufnimmt. Die Kopfverfolgungseinheit KVE arbeitet beispielsweise nach einem Verfahren, das in der Patentanmeldung mit dem amtlichen Aktenzeichen 19516664.7 beschrieben ist.

Bei dem dort beschriebenen Verfahren wird die räumliche Lage für ein aufgenommenes Zielobjekt, das während des dort beschriebenen Verfahrens klassifiziert wird, bestimmt. Dieses Verfahren wird bei dem

erfindungsgemäßen Verfahren dazu verwendet, daß der Kopf eines Betrachters B erkannt wird und die Bewegungen des Kopfes verfolgt werden. Somit ermittelt das in der Patentanmeldung 19516664.7 beschriebene Verfahren die räumliche Lage des Kopfes des Betrachters B des stereoskopischen Bildschirms DIS.

Verwendet wird in dem erfindungsgemäßen Verfahren jedoch nur die die horizontalen Bewegungen beschreibende Koordinate. Die beiden anderen Raumkoordinaten sind für das erfindungsgemäße Verfahren nicht von Interesse und werden aus diesem Grunde nicht benötigt.

In Fig. 2 ist das Verfahren zur Auswahl der zwei benötigten Kamerasignale Ki und Ki + 1 beschrieben.

Zuerst werden die beiden zulässigen Extrempositionen für die Kopfposition Pj des Betrachters B, also die Kopfposition P1 und die Kopfposition Pn-1 definiert.

Im Fall einer gewünschten naturgetreuen Übertragung, ist es vorteilhaft, die Kameras Kai äquidistant im Augenabstand aufzustellen, das heißt in diesem Fall wäre die Entfernung zwischen der Kopfposition P1 und der Kopfposition Pn-1 die gleiche, wie der Kameraabstand zwischen Kamera Ka1 und der Kamera Kan-1. Es werden dann in Abhängigkeit der Position des Betrachters B die nötigen Kamerasignale ausgewählt und übertragen.

Wenn der Kopf des Betrachters sich in der Kopfposition P1, also bei der in Fig. 1 beschriebenen Anordnung ganz links, oder sogar links davon befindet, werden von der Kopfverfolgungseinheit KVE die beiden Kamerasignale K1 und K2, also die Kamerasignale, die von den beiden links stehenden Kameras Ka1 und Ka2 erzeugten Kamerasignale, ausgewählt 2 und von der Kopfverfolgungseinheit KVE angefordert.

Es ist ebenso vorgesehen, daß nicht die Kamerasignale K1 und K2 ausgewählt werden, sondern es können auch beispielsweise die ganz links positionierte Kamera Ka1 und eine weitere Kamera Ka1 + x und die von diesen Kameras erzeugten Kamerasignale K1 und K1 + x ausgewählt werden. Die Auswahl der Kameras ist immer abhängig von dem Aufbau des stereoskopischen Bildschirms DIS.

Wenn sich der Kopf des Betrachters B auf der Kopfposition Pn-1 oder rechts davon befindet, ordnet die Kopfverfolgungseinheit KVE der Kopfposition Pn-1 die Kamerasignale Kn-1 und Kn zu, und fordert diese Kamerasignale von der Sendeeinheit SE an.

Es ist ebenso vorgesehen, daß nicht die Kamerasignale Kn-1 und Kn ausgewählt werden, sondern es können auch beispielsweise die ganz links positionierte Kamera Ka1 und eine weitere Kamera Ka1 + x und die von diesen Kameras erzeugten Kamerasignale Kn und Kn-x ausgewählt werden. Die Auswahl der Kameras ist auch in diesem Fall immer abhängig von dem Aufbau des stereoskopischen Bildschirms DIS.

Wenn sich der Kopf des Betrachters B zwischen den diskreten Kopfpositionen Pj befindet, wird jeweils die Kopfposition Pj des Betrachters B zugewiesen, die der tatsächlichen Kopfposition des Betrachters B am nächsten liegt.

Allgemein wird also durch die Kopfverfolgungseinheit KVE für eine Kopfposition Pj des Betrachters das Kamerasignal Ki als linkes Bild und das Kamerasignal Ki + x als rechtes Bild angeboten.

Es ist wichtig, die n-1 Kopfpositionen Pj des Betrachters B in Abhängigkeit des Kameraaufbaus zu definieren. Dies erfolgt in Abhängigkeit von dem Aufbau des stereoskopischen Bildschirms DIS und von den be-

absichtigten Effekten für den Betrachter B.

Dies bedeutet, daß von der Kopfverfolgungseinheit KVE zwei neue Kamerasignale angefordert werden. Um den stereoskopischen Effekt, das heißt den Tiefeneindruck für den Betrachter B zu verbessern bzw. zu übertreiben, ist es auch möglich, den Abstand der Kameras zu vergrößern.

Wenn der stereoskopische Effekt verringert werden soll, kann selbstverständlich auch der Abstand der Kameras verringert werden.

Nachdem die Kopfposition von der Kopfverfolgungseinheit KVE wie im vorigen beschrieben wurde, bestimmt ist 1, wählt die Kopfverfolgungseinheit KVE die benötigten zwei Ansichten, also die benötigten Kamerasignale  $K_i$  und  $K_{i+x}$  aus 2 und fordert diese von der Sendeeinheit SE an.

Dies geschieht in Form einer Kameraanforderungsnachricht KAN, die von der Empfängereinheit EE an die Sendeeinheit SE übertragen wird.

Nachdem die Sendeeinheit SE die Kameraanforderungsnachricht FAN empfangen hat, schaltet die Schaltereinheit S in der Weise, daß die beiden Eingänge der Schaltereinheit S an die angeforderten Kameras  $K_{ai}$  und  $K_{ai+x}$  gekoppelt werden, so daß die angeforderten Kamerasignale  $K_i$  und  $K_{i+x}$  an die stereoskopische Codiereinheit SC durchgeschaltet werden 3.

In der stereoskopischen Codiereinheit SC werden die durchgeschalteten Kamerasignale  $K_i$  und  $K_{i+x}$  zu einem zu sendenden codierten Videodatenstrom CV codiert 4.

Dieser codierte Videodatenstrom CV wird nun an die Empfängereinheit EE übertragen.

Der empfangene codierte Videodatenstrom CV wird in der Empfängereinheit EE in der stereoskopischen Decodiereinheit SD decodiert 5 und dem Betrachter B über den stereoskopischen Bildschirm DIS dargestellt 6.

In Fig. 3 ist eine vorteilhafte Verwendung des erfindungsgemäßen Verfahrens dargestellt.

Ein Problem bei dem Einsatz eines modellbasierten Codierungsverfahrens MC liegt in der Behandlung einer auftretenden Fehlersituation 8 bei den Codierungsverfahren MC, das heißt in einem Vorgehen, wenn das modellbasierte Codierungsverfahren Fehler aufweist, das heißt, wenn ein Bild von der modellbasierten Codierung MC falsch modelliert wurde.

Im Rahmen dieser Anmeldung ist unter dem modellbasierten Codierungsverfahren MC ein Verfahren zu verstehen, das die von den  $n$  Kameras aufgenommene Szene SZ durch drei-dimensionale Modelle darstellt. Dies ermöglicht eine weitere Datenreduktion der Bilddaten der aufgenommenen Szene SZ.

In diesem Fall ist eine Ausnahmebehandlung notwendig, die solange eine korrekte Übertragung des Videodatenstroms ermöglicht, bis wieder ein korrektes Modell für die Bilder des Videodatenstroms erzeugt, das heißt berechnet wurde. Für diese Ausnahmebehandlung bietet sich das im vorigen beschriebene Verfahren an.

Dies bedeutet, daß, wenn erkannt wird, daß ein Modell für einen Videodatenstrom CV mit schlechter Qualität, erstellt wurde, die modellbasierte Codierung MC des Videodatenstroms CV, die eine größere Kompression der Bilddaten ermöglicht, unterbrochen wird und das erfindungsgemäße Verfahren durchgeführt 10 wird, also daß die Szene jeweils von den zwei Kameras  $K_{ai}$  und  $K_{ai+1}$  aufgenommen wird, die durch die Kopfposition  $P_j$  des Betrachters B ausgewählt werden.

Wenn ein neues Modell, das die Bilddaten wieder korrekt beschreibt, erstellt wurde 9, kann wieder die mo-

dellbasierte Codierung MC durchgeführt werden 7.

Somit kann das erfindungsgemäße Verfahren im Zusammenhang mit der modellbasierten Codierung MC eines Videodatenstroms als Fallback-Lösung der modellbasierten Codierung MC verwendet werden.

#### Patentansprüche

1. Anordnung zur automatischen Auswahl zweier Kameras, die zwei Kamerasignale erzeugen, die zusammen eine stereoskopische Bildfolge bilden, aus einer Vielzahl von von  $n$  Kameras ( $K_{ai}; i=1 \dots n$ ) erzeugten Kamerasignalen ( $K_i; i=1 \dots n$ ).

— bei der in einer Sendeeinheit (SE) die  $n$  Kameras ( $K_{ai}$ ), die mehrere Ansichten einer Szene aufnehmen und einer Schaltereinheit (S) zur Verfügung stellen, in beliebiger Weise angeordnet sind;

— bei der in der Sendeeinheit (SE) eine stereoskopische Codiereinheit (SC) vorgesehen ist zur Codierung des zu sendenden codierten Videodatenstromes (CV),

— bei der in der Sendeeinheit (SE) die Schaltereinheit (S) vorgesehen ist, wobei zwei Ausgänge der Schaltereinheit (S) mit der stereoskopischen Codiereinheit (SC) gekoppelt sind und wobei  $n$  Eingänge der Schaltereinheit (S) mit den  $n$  Kameras ( $K_{ai} \dots K_{an}$ ) gekoppelt sind, die durch eine Kopfverfolgungseinheit (KVE) einer Empfängereinheit (EE) ausgewählte zwei Kamerasignale ( $K_i; K_{i+x}$ ) zur Verfügung stellen, wobei die eine Zahl ( $x$ ) eine beliebige Zahl im Bereich von 1 bis  $n$  ist,

— bei der in der Empfängereinheit (EE) die Kopfverfolgungseinheit (KVE) vorgesehen ist zur Bestimmung der Kopfposition ( $P_j; j=1 \dots n-1$ ), und zur Auswahl der der Kopfposition ( $P_j$ ) zugeordneten zwei Kamerasignale ( $K_i; K_{i+x}$ ),

— bei der in der Empfängereinheit (EE) eine stereoskopische Decodiereinheit (SD) vorgesehen ist zur Decodierung eines empfangenen codierten Videodatenstromes (CV), und

— bei der in der Empfängereinheit (EE) ein mit einem Ausgang der stereoskopischen Decodiereinheit (SD) gekoppelter stereoskopischer Bildschirm (DIS) vorgesehen ist.

2. Anordnung nach Anspruch 1, bei der der Abstand der  $n$  Kameras ( $K_{ai}$ ) äquidistant ist.

3. Verfahren zur automatischen Auswahl zweier Kameras ( $K_{ai}; K_{ai+x}$ ), die zwei Kamerasignale ( $K_i; K_{i+x}$ ) erzeugen, die zusammen eine stereoskopische Bildfolge bilden, aus einer Vielzahl von von  $n$  Kameras ( $K_{ai}; i=1 \dots n$ ) erzeugten Kamerasignalen ( $K_i; i=1 \dots n$ ),

— bei dem in einer Empfängereinheit (EE) durch eine Kopfverfolgungseinheit (KVE) eine Kopfposition ( $P_j; j=1 \dots n-1$ ) eines Betrachters (B) bestimmt wird,

— bei dem von der Kopfverfolgungseinheit (KVE) der Kopfposition ( $P_j$ ) zugeordnete zwei von zwei Kameras ( $K_{ai}; K_{ai+x}$ ) erzeugten Kamerasignale ( $K_i; K_{i+1}$ ) ausgewählt werden, wobei der Kopfposition ( $P_j$ ) die zwei Kamerasignale ( $K_i$ ) und ( $K_{i+x}$ ) zugeordnet werden, wobei eine Zahl ( $x$ ) eine beliebige Zahl im Bereich von 1 bis  $n$  ist,

— bei dem die Empfängereinheit (EE) eine Ka-

mera Anforderungsnachricht (KAN) an eine Sendeeinheit (SE) sendet, mit der die zwei benötigten Kamerasignale ( $K_i$ ;  $K_i + x$ ) angefordert werden,

— bei dem die Sendeeinheit (SE) die Kameraanforderungsnachricht (KAN) empfängt und speichert,

— bei dem in der Sendeeinheit (SE) durch eine Schaltereinheit (S) die durch die Kopfverfolgungseinheit (KVE) ausgewählten zwei Kamerasignale ( $K_i$ ;  $K_i + 1$ ) an eine stereoskopische Codiereinheit (SC) durchgeschaltet werden,

— bei dem in der Sendeeinheit (SE) in einer stereoskopischen Codiereinheit (SC) ein zu sendender codierter Videodatenstrom (CV) codiert wird,

— bei dem der zu sendende codierte Videodatenstrom (CV) von der Sendeeinheit (SE) an die Empfängereinheit (EE) übertragen wird,

— bei dem der empfangene codierte Videodatenstrom (CV) in der Empfängereinheit (EE) gespeichert wird, und

— bei dem in der Empfängereinheit (EE) der empfangene codierte Videodatenstrom (CV) in einer stereoskopischen Decodiereinheit (SD) decodiert wird.

4. Verwendung des Verfahrens nach Anspruch 3 bei Durchführung eines modellbasierten Codierungsverfahrens (MC) zur Codierung des Videodatenstromes (CV), wenn die modellbasierte Codierung (MC) ausfällt, solange bis die modellbasierte Codierung (MC) wieder eingesetzt werden kann.

Hierzu 3 Seite(n) Zeichnungen

35

40

45

50

55

60

65

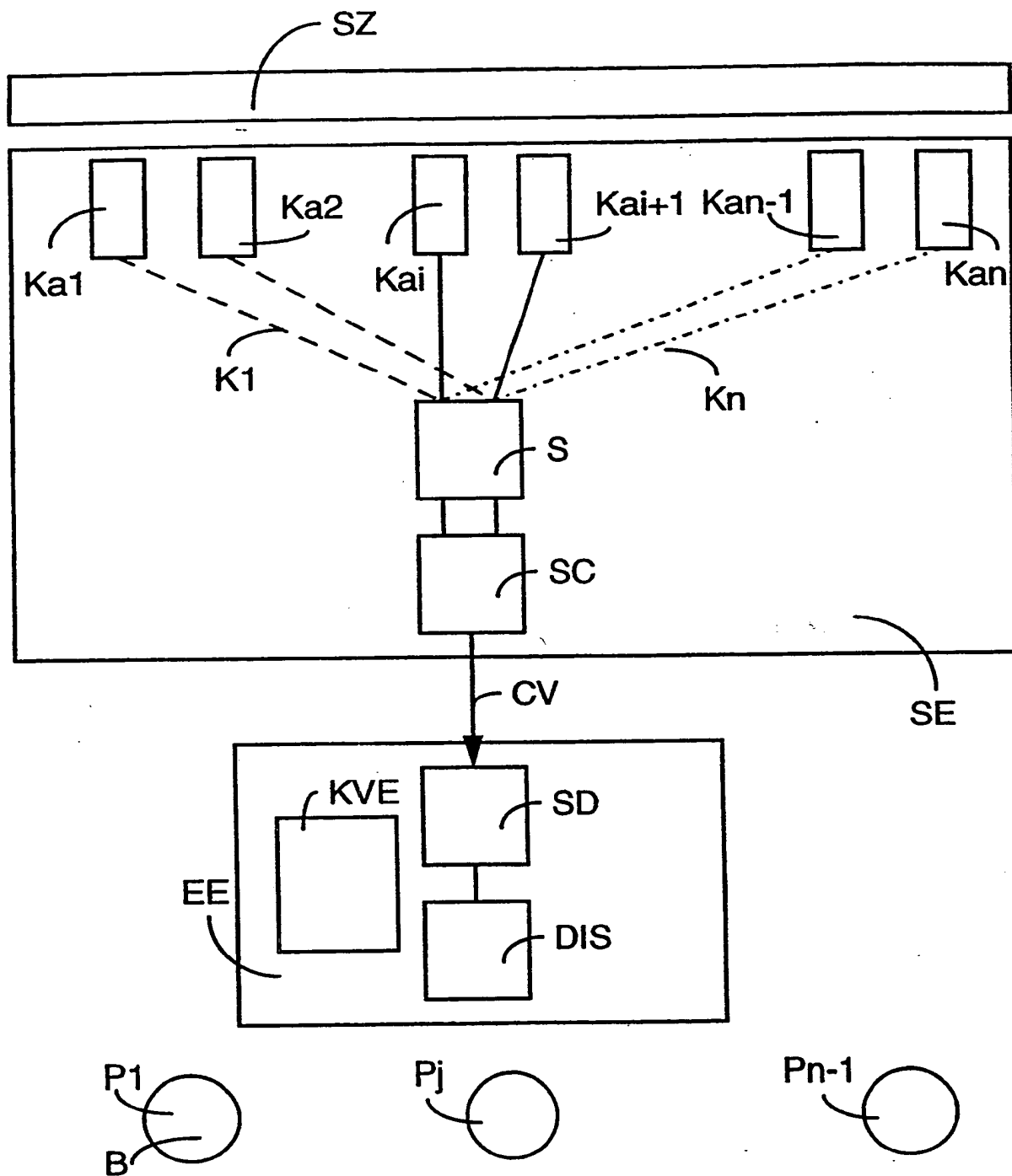


Fig. 1

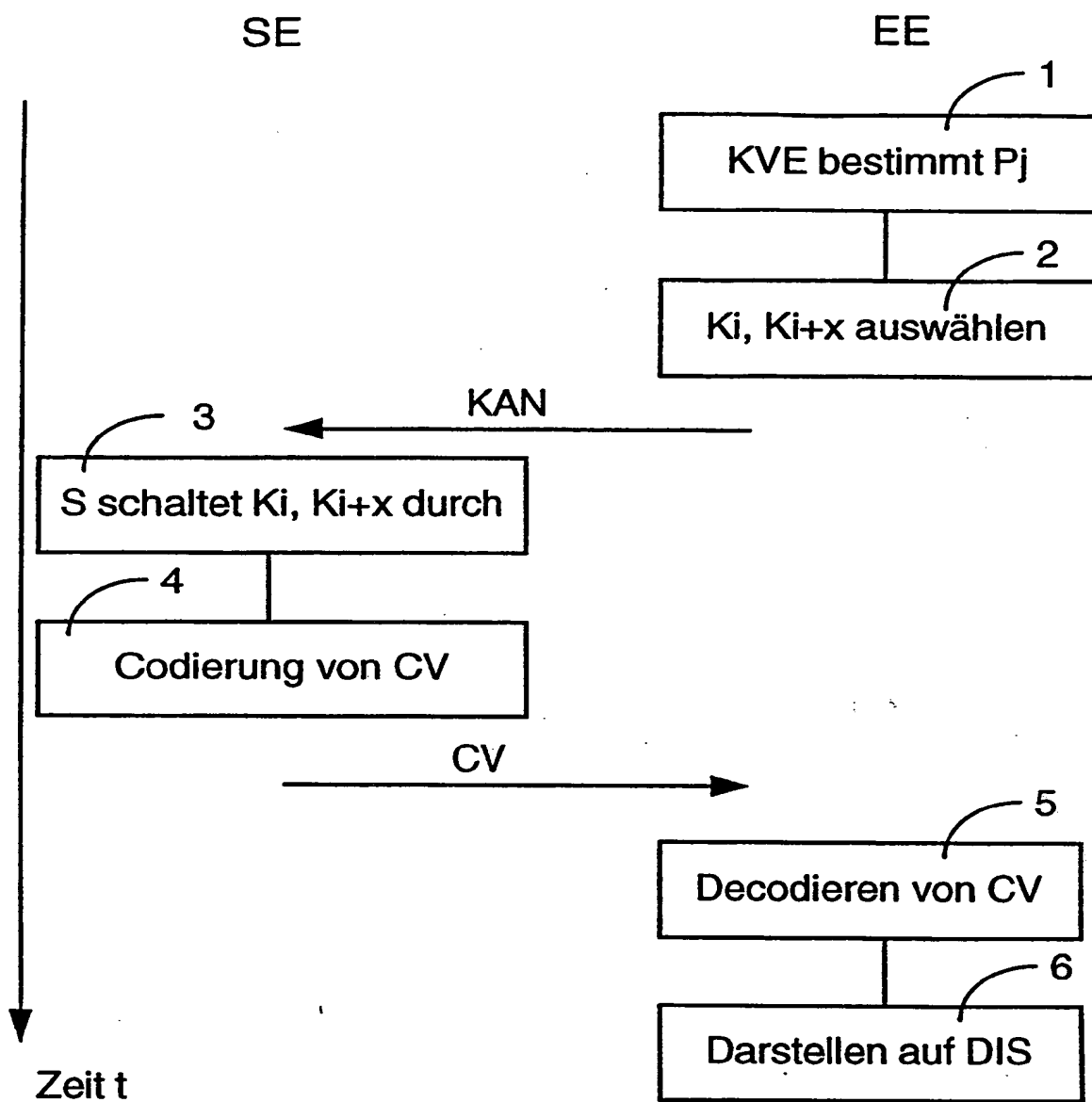


Fig. 2

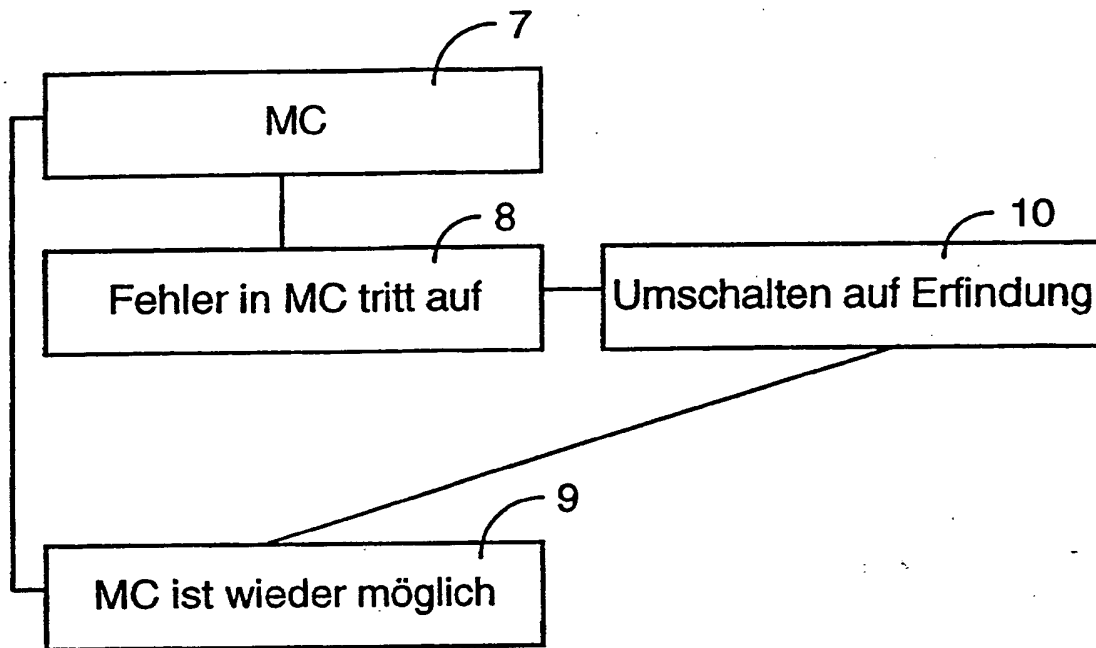


Fig. 3



CFP 1239 EP

- 1 -

- (19) Federal Republic of Germany - German Patent Office  
(12) Patent Specification  
(10) DE 195 28 425 C1  
(51) Int. Cl.: H 04 N 13/00  
(21) Application Number: 195 28 425.9-31  
(22) Date of Filing: 2<sup>nd</sup> August 1995  
(43) Date of Publication: --  
(45) Date of Publication of Issue of Patent: 15<sup>th</sup> May 1996

Appeals may be lodged within three months of publication

---

(73) Assignee:

Siemens AG, 80333 Munich, Germany

(72) Inventors:

Riegel, Thomas, Dipl.-Inform, 81739 Munich, Germany;

Ziegler, Manfred, Dipl.-Inform, 81739 Munich, Germany

(56) For the assessment of patentability the following documents were taken into consideration:

DE 40 27 471 C2

EP 06 50 301 A2

---

(54) Configuration and procedure for the automatic selection of two camera signals from multiple camera signals by determining the position of the head of a viewer.

**THIS PAGE BLANK (USPTO)**

(57). A head position ( $P_j$ ) of a viewer (B) is determined by a head-following unit (KVE), and two camera signals ( $K_i$ ;  $K_i + x$ ) are selected, and requested from a transmitter unit (SE). The requested cameras ( $K_{ai}$ ;  $K_{ai} + x$ ) are switched through to a stereoscopic encoding unit (SC) by a switching unit (S), where they are encoded into an encoded video data stream (CV). This encoded video data stream (CV) is transmitted to a receiver unit (EE), where it is decoded and presented to the viewer (B) on a stereoscopic screen (DIS).

In this way, a very simple, automated selection of two necessary camera signals ( $K_i$ ;  $K_i + x$ ) out of multiple camera signals ( $K_i$ ) is achieved.

**THIS PAGE BLANK (USPTO)**

### Description

During the transmission of so-called multiview stereo picture sequences, in which more than two cameras record a scene, several problems exist during the transmission of the camera signals that describe the recorded scene. With multiview stereo picture sequences, the receiver needs two camera signals that describe the same scene from different views. A depth effect is achieved for the viewer through the existence of a disparity between the pictures of the recorded camera signals, that is to say, through the local displacement of corresponding pixels between the pictures of the two camera signals, which are presented via a receiver on a stereoscopic screen. In other words, the viewer has the impression that he or she is observing a three-dimensional picture.

If a scene is recorded by several cameras, all camera signals were transmitted by all cameras to the receiver, and the selection of the two necessary camera signals were to take place in the receiver, then an extremely large transmission capacity would be required, which would include much unnecessary redundancy.

As only two camera signals are needed in the receiver unit (the two 'correct' ones for the viewer), it is desirable to actually transmit only the necessary two camera signals. Furthermore, it is desirable to be able to

**THIS PAGE BLANK (USPTO)**

vary the view of the scene without driving up the cost.

The use of more than two cameras for stereoscopic picture recording is hitherto known from EP 0650301 A2. Furthermore, DE 40 27 471 C2 concerns a spatial representation of pictures that depends on the detected position of the viewer.

Therefore, the object of the invention is to solve the problem of specifying a configuration and a procedure by which two camera signals are chosen in a simple manner from camera signals produced by a plurality of  $n$  cameras which record a scene from different views, the signals being chosen automatically by a viewer of the scene, who is located at the receiver.

This problem is solved by the configuration according to Claim 1, and the procedure according to Claim 3.

One advantage of the configuration according to the present invention lies, above all, in the fact that the configuration is very simple and therefore only minor enhancements to configurations are necessary, which are in any case required in order to realize a stereoscopic video data stream on the basis of two connected cameras which record a scene.

Also, the procedure according to the present invention is characterized by its simplicity, which leads to a very fast execution of the procedure, whereby it is possible to

**THIS PAGE BLANK (USPTO)**



implement the procedure in real time, without delay effects, which may be disturbing for the viewer, when selecting another view of an observed scene.

The use of the procedure according to Claim 4 for exception handling in case of a mistake, and therefore as a so-called fallback solution, in the use of a model-based encoding procedure for the encoding of a video data stream, significantly increases the dependability of the model-based encoding procedure.

Further refinements of the invention have resulted from the attached claims.

A preferred embodiment of the invention and an advantageous application of the procedure according to the invention are presented in the drawings and are more closely described in the following.

The following are shown:

Fig. 1 is a diagram that shows the configuration of a receiver unit, a transmitter unit, and multiple cameras, which are selected based on the position of the viewer's head;

Fig. 2 is a flow chart that describes the procedure according to the invention in accordance with Claim 3; and

Fig. 3 is a flow chart that shows the use of the procedure according to the invention as a fallback solution in model-based encoding.

**THIS PAGE BLANK (USPTO)**

Based on Figs. 1 to 3, the configuration according to the invention, the procedure according to the invention, and an application of the procedure according to the invention are further illustrated.

In Fig. 1, the configuration according to the invention is described, which features at least the following components:

- a transmitter unit SE which features at least the following components:
  - a plurality of cameras Kai, which record a scene from several views, in other words from different perspectives; a first index, i, which unambiguously identifies each camera, is a number that lies in the range between 1 and n, the values 1 and n being contained in the range;
  - a switching unit S, with n inputs, wherein each input of switching unit S is connected to a separate camera Kai, and two outputs,
  - a stereoscopic encoding unit SC for encoding a video data stream CV to be transmitted;
- a receiver unit EE which features at least the following components:
  - a stereoscopic decoder unit SD, in which the received encoded video data stream CV is decoded,
  - a head-following unit KVE, in which the head

**THIS PAGE BLANK (USPTO)**

position  $P_j$  of a viewer B is determined, and in which the necessary two views are selected from the determined coordinates of the head position, in other words, two necessary camera signals  $K_i$  are selected, wherein, as well as the first index  $i$  denoting the camera signal provided by the  $i$ -th camera, a second index  $j$ , which unambiguously identifies each head position  $P_j$  of the viewer B, is a number lying in the range between 1 and  $n - 1$ , the values 1 and  $n - 1$  being contained in the range;

- a stereoscopic screen DIS, which presents a decoded video data stream DCV to the viewer B in such a way as to create a stereoscopic effect for the viewer B, that is to say, an impression of depth is formed as the picture is watched.

The cameras  $K_{ai}$  record a scene SZ. The cameras  $K_{ai}$  are arranged in any desired manner.

This is a prerequisite for the fact that in any desired selection of two cameras  $K_{ai}$  and  $K_{ai + x}$ , in which the number  $x$  lies in the range between 1 and  $n - 1$ , a stereoscopic effect can be achieved for the viewer B.

The transmitter unit SE features at least the  $n$  cameras  $K_{ai}$ , the switching unit S, and the stereoscopic encoding unit SC.

**THIS PAGE BLANK (USPTO)**

The switching unit S is connected by its two outputs to the stereoscopic encoding unit SC. The design of the stereoscopic encoding unit SC is well-known to one skilled in the art and is, for example, described in (R. ter Horst, A Demonstrator for Coding and Transmission of Stereoscopic Video Signals, Proceedings of 4th European Workshop on Three-dimensional Television, Rome, 20-21 October 1993, P. 273-280, 1993).

The video data stream which is to be transmitted, and which is encoded in the stereoscopic encoding unit SC, is transmitted to the receiver unit EE. Each of the inputs of the switching unit S is connected to a separate camera Kai.

Corresponding to a camera request message KAN, which will be further described, the cameras Kai and Kai + 1, which are requested by the camera request message KAN, are selected in the switching unit S. In this way, each of the requested camera signals Ki and Ki + 1, from the cameras Kai and Kai + 1, are encoded and transmitted.

Thus in each case only the requested camera signals are selected and switched through 3 to the stereoscopic encoding unit SC, where they are encoded 4 into the video data stream CV and transmitted to the receiver unit EE.

In the receiver unit EE, the encoded video data stream CV is received and decoded 5. Subsequently, the decoded video data stream is presented on the stereoscopic screen

**THIS PAGE BLANK (USPTO)**



DIS to the viewer B.

In 6, the receiver unit EE features at least the stereoscopic decoder unit SD, the head-following unit KVE, and the stereoscopic screen DIS.

The stereoscopic decoder unit SD is connected to the stereoscopic screen DIS. In the head-following unit KVE, the position of the head of the viewer B is determined by means of a further camera, with which the head of the viewer B is recorded and detected, and the spatial location of the head of the viewer B, that is the head position  $P_j$  of the viewer B, is determined.

Based on the head position  $P_j$ , wherein the second index  $j$  lies in the range between 1 and  $n - 1$ , the camera signals  $K_i$  and  $K_i + 1$ , which are assigned to the head position  $P_j$ , are selected and requested by the head-following unit KVE.

In the head-following unit KVE, the head position  $P_j$  of the viewer B is determined 1 and the horizontal movements of the viewer B with respect to the screen are followed. For this, at least one camera that records the head of the viewer B is used. The head-following unit KVE may operate, for example, according to the procedure described in the patent application with application number 19516664.7.

In the procedure described there, the spatial location

**THIS PAGE BLANK (USPTO)**

is determined for a recorded target object, which is classified during the procedure described there. This procedure is used in the procedure according to the present invention so that the head of the viewer B may be detected and the movements of the head followed. In this way the procedure described in patent application 19516664.7 determines the spatial location of the head of the viewer B of the stereoscopic screen DIS.

However, only those coordinates that describe horizontal movements are used in the procedure according to the present invention. Neither of the other spatial coordinates are of interest for the procedure according to the invention and for this reason are not required.

In Fig. 2, the procedure in which the two necessary camera signals  $K_i$  and  $K_i + 1$  are selected is described.

First of all, both of the acceptable extreme positions for the head position  $P_j$  of the viewer B are defined, that is to say the head position  $P_1$  and the head position  $P_n - 1$ .

In the case that a realistic transmission is desired, it is advantageous to set up the cameras  $K_{ai}$  equidistant from each eye, which in this case means that the distance between the head position  $P_1$  and the head position  $P_n - 1$  would be the same as the distance between the camera  $K_{a1}$  and the camera  $K_{an} - 1$ . The necessary camera signals can then be selected and transmitted, depending on the position

**THIS PAGE BLANK (USPTO)**

of the viewer B.

If the head of the viewer is located at head position P1, in other words on the far left of the configuration described in Fig. 1, or even further to the left, both of the camera signals K1 and K2 are selected, that is to say the camera signals from the two cameras positioned at the left Ka1 and Ka2. These signals are then requested by the head-following unit KVE.

This set-up also provides for a situation in which camera signals K1 and K2 are not selected, but rather there could for example be a camera positioned on the far left Ka1 and a further camera,  $Ka1 + x$ , and the camera signals produced by these cameras K1 and  $K1 + x$  are selected. The selection of the cameras is always dependent on the configuration of the stereoscopic screen DIS.

If the head of the viewer B is located at head position  $Pn - 1$  or to the right of it, the head-following unit KVE assigns the camera signals  $Kn - 1$  and  $Kn$  to the head position  $Pn - 1$ , and requests these camera signals from the transmitter unit SE.

The set-up also provides for a situation in which camera signals  $Kn - 1$  and  $Kn$  are not selected, but rather there could for example be a camera positioned on the far left Ka1 and a further camera,  $Ka1 + x$ , and the camera signals produced by these cameras  $Kn$  and  $Kn - x$  are

**THIS PAGE BLANK (USPTO)**

selected. In this case too, the selection of the cameras is always dependent on the configuration of the stereoscopic screen DIS.

If the head of the viewer B is located between the discrete head positions  $P_j$ , the head position  $P_j$  of the viewer B that is assigned is the one that is closest to the actual head position of the viewer B.

In general, therefore, the camera signal  $K_i$  as a left picture and the camera signal  $K_i + x$  as a right picture for a head position  $P_j$  of the viewer are offered through the head-following unit KVE.

It is important to define the  $n - 1$  head positions  $P_j$  of the viewer B independently from the camera arrangement. This is carried out depending on the structure of the stereoscopic screen DIS and from the intended effects for the viewer B.

This means that two new camera signals are requested by the head-following unit KVE. In order to improve and/or exaggerate the stereoscopic effect, that is the impression of depth for the viewer B, it is also possible to increase the space between the cameras.

If the stereoscopic effect is supposed to be reduced, naturally the space between the cameras can also be reduced.

After the head position is determined 1 by the head-following unit KVE, as previously described, the head-

**THIS PAGE BLANK (USPTO)**



following unit KVE selects 2 the necessary two views, that is to say, the necessary camera signals  $K_i$  and  $K_i + x$ , and requests these from the transmitter unit SE.

This happens in form of a camera request message KAN that is transmitted by the receiver unit EE to the transmitter unit SE.

After the transmitter unit SE has received the camera request message FAN, the switching unit S switches in such a way that both inputs of the switching unit S are connected to the requested cameras  $K_{ai}$  and  $K_{ai} + x$ , so that the requested camera signals  $K_i$  and  $K_i + x$  are switched through 3 to the stereoscopic encoding unit SC.

In the stereoscopic encoding unit SC, the switched through camera signals  $K_i$  and  $K_i + x$  are encoded 4 into an encoded video data stream CV to be sent.

This encoded video data stream CV is then transmitted to the receiver unit EE.

The received encoded video data stream CV is decoded in the stereoscopic decoder unit SD 5 within the receiver unit EE, and is presented 6 to the viewer B on the stereoscopic screen DIS.

In Fig. 3, an advantageous application of the procedure according to the invention is shown.

One problem in the deployment of a model-based encoding procedure MC lies in the way in which an error

**THIS PAGE BLANK (USPTO)**

situation 8 in the encoding procedure MC is handled when it occurs. In practice this means when an error occurs in the model-based encoding procedure, that is to say, when a picture was incorrectly modeled by the model-based encoding MC.

In the scope of this application, under the model-based encoding procedure MC, a procedure in which the scene SZ, having been recorded by the n cameras, is presented through three-dimensional models is to be understood. This enables a further reduction in the picture data of the recorded scene SZ.

In this case, exception handling that enables correct transmission of the video data stream for as long as is required for a correct model for the pictures of the video data stream to be produced, that is to say, calculated is necessary. The previously described procedure can be used for this exception handling.

This means that, if it is known that a model was created for a bad quality video data stream CV, the model-based encoding MC of the video data stream CV, which enables a larger compression of the picture data, is interrupted, and the procedure according to the invention is carried out 10. So the scene is recorded as appropriate by the two cameras Kai and Kai + 1, which are selected by the head position Pj of the viewer B.

**THIS PAGE BLANK (USPTO)**

If a new model that once more correctly describes the picture data were created 9, the model-based encoding MC can be carried out again 7.

In this way, the procedure according to the invention can be used in conjunction with the model-based encoding MC of a video data stream as a fallback solution of the model-based encoding MC.

#### Claims

1. A configuration for the automatic selection of two cameras, which produce two camera signals that together form a stereoscopic picture sequence, out of a plurality of camera signals ( $K_i$ ;  $i = 1 \dots n$ ) produced by  $n$  cameras ( $K_{ai}$ ;  $i = 1 \dots n$ ):

- wherein, in a transmitter unit (SE), the  $n$  cameras ( $K_{ai}$ ), which record several views of a scene, as well as one switching unit (S) are available, and are arranged in any desired manner;
- wherein, in the transmitter unit (SE), a stereoscopic encoding unit (SC) is provided for encoding an encoded video data stream (CV) to be sent;
- wherein, in the transmitter unit (SE), the switching unit (S) is provided, wherein two outputs of the switching unit (S) are connected to the stereoscopic

**THIS PAGE BLANK (USPTO)**

encoding unit (SC), and wherein  $n$  inputs of the switching unit (S) are connected to the  $n$  cameras ( $K_{ai}$  . . .  $K_{an}$ ), which provide two camera signals ( $K_i$ ;  $K_i + x$ ) that are selected by a head-following unit (KVE) of a receiver unit (EE), the one number ( $x$ ) being any number in the range between 1 and  $n$ ;

- wherein, in the receiver unit (EE), the head-following unit (KVE) is provided to determine the head position ( $P_j$ ;  $j = 1 \dots n - 1$ ), and to select the two camera signals ( $K_i$ ;  $K_i + x$ ) which are assigned to the head position ( $P_j$ );

- wherein, in the receiver unit (EE), a stereoscopic decoder unit (SD) is provided for decoding the received encoded video data stream (CV), and

- wherein, in the receiver unit (EE), a stereoscopic screen (DIS), which is connected to the output of the stereoscopic decoder unit (SD), is provided.

2. The configuration according to claim 1, wherein the distance between the  $n$  cameras ( $K_{ai}$ ) is uniform.

3. A procedure for the automatic selection of two cameras ( $K_{ai}$ ;  $K_{ai} + x$ ), which produce two camera signals ( $K_i$ ;  $K_i + x$ ), which together form a stereoscopic picture sequence, out of a plurality of camera signals ( $K_i$ ;  $i = 1 \dots n$ ) produced by  $n$  cameras ( $K_{ai}$ ;  $i = 1 \dots n$ ):

- wherein, in a receiver unit (EE), a head position

**THIS PAGE BLANK (USPTO)**



( $P_j$ ;  $j = 1 \dots n - 1$ ) of a viewer (B) is determined by a head-following unit (KVE);

- wherein, two camera signals ( $K_i$ ,  $K_i + 1$ ) which are produced by two cameras ( $K_{ai}$ ;  $K_{ai} + x$ ) that are assigned to the head position ( $P_j$ ) by the head-following unit (KVE) are selected, wherein a number ( $x$ ) is any desired number in the range between 1 and  $n$ ;

- wherein, the receiver unit (EE) sends a camera request message (KAN) to a transmitter unit (SE), by which the two necessary camera signals ( $K_i$ ;  $K_i + x$ ) are requested;

- wherein, the transmitter unit (SE) receives and stores the camera request message (KAN);

- wherein, in the transmitter unit (SE), the two camera signals ( $K_i$ ;  $K_i + 1$ ) which are selected by the head-following unit (KVE) are switched through to a stereoscopic encoding unit (SC) through a switching unit (S);

- wherein, in the transmitter unit (SE), an encoded video data stream (CV) to be sent is encoded in a stereoscopic encoding unit (SC);

- wherein, the encoded video data stream (CV) to be sent is transmitted from the transmitter unit (SE) to the receiver unit (EE);

- wherein, the received encoded video data stream (CV) is stored in the receiver unit (EE); and

**THIS PAGE BLANK (USPTO)**

- wherein, in the receiver unit (EE), the received encoded video data stream (CV) is decoded in a stereoscopic decoder unit (SD).

4. An application of the procedure according to Claim 3, wherein, in the execution of a model-based encoding procedure (MC) to encode the video data stream (CV), if the model-based encoding (MC) drops out, the procedure is executed for as long as is required until the model-based encoding (MC) can once more be implemented.

Three pages of drawings attached

**THIS PAGE BLANK (USPTO)**

Fig. 2

1. KVE determines  $P_j$
2.  $K_i$ ,  $K_i + x$  selection
3. S switches through  $K_i$ ,  $K_i + x$
4. Encoding of CV
5. Decoding of CV
6. Presentation on DIS

Time t

Fig. 3

7. MC
8. Mistake in MC occurs
9. MC is possible once again
10. Switch to invention

**THIS PAGE BLANK (USPTO)**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**